IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of	Atty. Docket
Fabio VIGNOLI	NL 021053
	Confirmation No. 1612
Serial No. 10/532,469	Group Art Unit: 2626
Filed: APRIL 22, 2005	Examiner: SAINT CYR, Leonard

Title: CONTROLLING AN APPARATUS BASED ON SPEECH

Mail Stop Appeal Brief-Patents
Board of Patent Appeals and Interferences
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF

Sir:

Appellant herewith respectfully presents a Brief on Appeal as follows, where a Notice of Appeal is filed concurrently having paid the Notice of Appeal fee on April 21, 2009:

REAL PARTY IN INTEREST

The real party in interest in this appeal is the assignee of record Koninklijke Philips Electronics N.V., a corporation of The Netherlands having an office and a place of business at Groenewoudseweg 1, Eindhoven, Netherlands 5621 BA.

RELATED APPEALS AND INTERFERENCES

Appellant and the undersigned attorney are not aware of any other appeals or interferences which will directly affect or be directly affected by or having a bearing on the Board's decision in the pending appeal.

STATUS OF CLAIMS

Claims 1-20 are pending in this application. Claims 1-20 are rejected in the Final Office Action mailed in May 12, 2010. Claims 1-20 are the subject of this appeal.

STATUS OF AMENDMENTS

Appellant did not file a response to a Final Office Action mailed on May 12, 2010. This Appeal Brief is in response to the Final Office Action mailed May 12, 2010, that finally rejected claims 1-20.

SUMMARY OF THE CLAIMED SUBJECT MATTER

The present invention, for example, as recited in independent claim 1 and shown in FIGs 1-2, as well as described on page 6, lines 1-6 of the specification, is directed to a speech control unit 100 for controlling an apparatus 200 on basis of speech, comprising a microphone array, comprising multiple microphones 102, 104, 106, 108, 110 for receiving respective audio signals 103, 105, 107, 109, 111.

As shown in FIG 1, and described on page 6, lines 7-9 and page 7, lines 8-18, the speech control unit 100 includes a beam forming module 116 for extracting a speech signal 117 of a user U1, from the audio signals 103, 105, 107, 109, 111 as received by the microphones 102, 104, 106, 108, 110, by enhancing first components of the audio signals which represent an utterance originating from a first position P1 of the user U1 relative to the microphone array.

As shown in FIG 1, and described on page 6, lines 13-14 and page 7, lines 8-18, a speech recognition unit 118 creates an instruction for the apparatus 200 based on recognized speech items of the speech signal. Further, as shown in FIG 1, and described on page 6, lines 10-12 and page 7, lines 8-18, a keyword recognition system 120 recognizes a predetermined keyword that is spoken by the user U1 and which is represented by a particular audio signal.

As described on page 3, lines 22-29, and page 6, line 10 to page 7, line 3 of the

specification, the speech control unit 100 is arranged to control the beam forming module 116, on basis of the recognition of the predetermined keyword, in order to enhance second components of the audio signals which represent a subsequent utterance originating from a second position P2 of the user U1 relative to the microphone array.

Further, as described on page 4, lines 4-15 of the specification, the recognition of the predetermined keyword at the second position calibrates the beam forming module 116 to follow the user from the first position P1 to the second position P2 so that the subsequent utterance originating from the second position P2 are accepted while utterances of other users U2 at other positions are discarded. The second position P2 includes orientation and distance relative to the microphone array, as described on page 4, lines 22-25.

In addition, as described on page 4, lines 25-28, the speech control unit 100 is also configured to discriminate between sounds originating from users who are located in front of each other relative the microphone array. The subsequent utterance originating from the second position P2 will be discarded if not preceded by the recognition of the predetermined keyword originating from the second position P2, as described on page 3, line 29 to page 4, line 3.

The present invention, for example, as recited in independent claim 9 and shown in

FIGs 1-2, as well as described on page 6, lines 1-6 of the specification, is directed to a method of controlling an apparatus 200 on basis of speech, comprising receiving respective audio signals 103, 105, 107, 109, 111 by a microphone array, comprising multiple microphones 102, 104, 106, 108, 110.

As shown in FIG 1, and described on page 6, lines 7-14 and page 7, lines 8-18, the method further includes extracting a speech signal 117 of a user U1, from the audio signals 103, 105, 107, 109, 111 as received by the microphones 102, 104, 106, 108, 110, by enhancing first components of the audio signals which represent an utterance originating from a first position P1 of the user relative to the microphone array.

Further, as shown in FIG 1, and described on page 6, lines 10-12 and page 7, lines 8-18 the method include recognizing a predetermined keyword that is spoken by the user U1 based on a particular audio signal.

In addition, as described on page 3, lines 22-29, page 6, line 10 to page 7, line 3, and page 4, lines 4-15 of the specification, the extraction of the speech signal of the user U1 is controlled on basis of the recognition of the predetermined keyword, in order to enhance second components of the audio signals which represent a subsequent utterance originating from a second position P2 of the user U1 relative to the microphone array, while discarding utterances of other users at other positions.

Further, as described on page 4, lines 22-28 of the specification, the second position P2 includes an orientation and a distance relative to the microphone array so that sounds

originating from users who are located in front of each other, relative the microphone array, are discriminated.

As described on page 3, line 29 to page 4, line 3, an instruction for the apparatus is created based on recognized speech items of the speech signal, and the subsequent utterance originating from the second position P2 is discarded if not preceded by the recognition of the predetermined keyword originating from the second position P2.

The present invention, for example, as recited in dependent claim 20 and shown in FIG 1, as well as described on page 6, lines 10-12 and page 7, lines 8-18 of the specification, is directed to having the beam forming module 118 being connected to the microphone array, and the keyword recognition system 120 is connected to one microphone 110 of the microphone array for detecting the predetermined keyword. As clearly shown in FIG 1, the keyword recognition system 120 is further connected to the beam forming module 116 for providing the detected predetermined keyword to the beam forming module 116.

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Whether claims 1-10, 15 and 20 of U.S. Patent Application Serial No. 10/532,469 are unpatentable under 35 U.S.C. §103(a) over U.S. Patent No. 7,136,817 (Schroder) in view of U.S. Patent No. 7,050,971 (Kaufholz) and U.S. Patent Application Publication No. 2002/0150263 (Rajan).

Whether claims 11-14 and 16-19 of U.S. Patent Application Serial No. 10/532,469 are unpatentable under 35 U.S.C. §103(a) over Schroder in view of Kaufholz, Rajan and U.S. Patent Application Publication No. 2002/0181723 (Kataoka).

ARGUMENT

Claims 1-10, 15 and 20 are said to be unpatentable over Schroder, Kaufholz and Rajan.

Appellant respectfully requests the Board to address the patentability of independent claims 1 and 9, dependent claim 20, and further claims 2-8 and 10-19 as depending from independent claims 1 and 9, based on the requirements of independent claims 1 and 9. This position is provided for the specific and stated purpose of simplifying the current issues on appeal. However, Appellant herein specifically reserves the right to argue and address the patentability of claims 2-8 and 10-19 at a later date should the separately patentable subject matter of claims 2-8 and 10-19 later become an issue. Accordingly, this limitation of the subject matter presented for appeal herein, specifically limited to discussions of the patentability of independent claims 1 and 9 is not intended as a waiver of Appellant's right to argue the patentability of the further claims and claim elements at that later time.

Schroder is directed to voice control of a device where voices of users are characterized to distinguish or identify the speech inputs of different users. This enables the device to recognize and respond only to a particular user. Schroder uses user **profiles** to distinguish or identify the speech inputs of different users.

On page 5, second full paragraph of the Final Office Action, column 2, lines 39-44,

column 3, lines 49-52, and column 1, lines 44-47 of Schroder are cited to allegedly show that a "subsequent utterance originating from the second position will be discarded if not preceded by the recognition of the predetermined keyword originating **from the second position**," as recited in independent claim 1, and similarly recited in independent claim 9. (Illustrative emphasis provided)

This allegation is respectfully traversed. In particular, the noted sections of Schroder specifically recite the following, where column 2, lines 39-44 recites:

> Similarly, an operator-control command which, after its input by the first user, allows voice commands from a second user to be accepted may be advantageously provided. This makes it possible to pass on operator-control authority in a way corresponding to the passing on of a remote control unit from a first user to a second user. (Emphasis added)

Column 3, lines 49-52 recites:

> If this is the case, the input command for controlling the voice-controlled system is used in method step 8, for example for menu control or navigation. (Emphasis added)

Column 1, lines 44-47 recites:

> Convenience can be enhanced if, for speech input, one or more microphones are provided in the device appertaining to consumer electronics, so that the user can carry out operator control from any desired place in the room without taking along the remote control unit. (Emphasis added)

The above-noted portions of Schroder merely refer to controlling a device by voice using microphones and passing the control from one user to another. Such a disclosure has nothing to do and does not disclose or suggest with discarding anything, let alone

discarding a subsequent utterance from the second position, and doing so if the subsequent utterance is not preceded by the recognition of the predetermined keyword originating from the second position, as recited in independent claims 1 and 9.

Even assuming, arguendo, that the noted portions of Schroder have something to do with discarding, which they do not, there is still no disclosure or suggestion of discarding an utterance "originating **from the second position** ... if not preceded by the recognition of the predetermined keyword originating **from the second position**," as recited in independent claim 1, and similarly recited in independent claim 9. (Illustrative emphasis provided)

Even further assuming, arguendo, that these features are somehow disclosed or suggested in Schroder, page 5, last paragraph of the Office Action correctly notes that Schroder and Kaufholz do not disclose or suggest the following features recited in independent claim 1, and similarly recited in independent claim 9 (illustrative emphasis provided):

> wherein the recognition of the predetermined keyword at the second position calibrates the beam forming module to follow the user from the first position to the second position so that the subsequent utterance originating from the second position are accepted while **utterances of other users at other positions are discarded**, the second position including an orientation and a **distance relative to the microphone array**, and the speech control unit being configured to discriminate between sounds originating from users who are located **in front of each other relative the microphone array**.

Rajan is cited in an attempt to remedy the deficiencies in Schroder and Kaufholz.

Rajan is directed to a signal processing system which receives signals from different users and sensors. The Rajan system **separates** signals from each of the sources or users 1-1, 1-2, 1-3 in a meeting, as shown in FIG 1 and described in paragraph [0022]. It is respectfully submitted that separating or isolating signals does not disclose or suggest discarding any signal. While Rajan discloses to separate signals from different users/sources, there is no disclosure or suggestion in Rajan that "utterances of **other** users at **other** positions are **discarded**," as recited in independent claim 1, and similarly recited in independent claim 9. (Illustrative emphasis provided)

Further, paragraph [0057], last eight lines of Rajan recite:

> The predetermined curved plots used may be circular arcs, in which case, the spectrogram processing module 33 will be able to estimate, not only the direction from which the speech emanated, but also the distance from the microphones of that user (since it would be able to determine the centre of the circle corresponding to the circular arc which fits the determined time delay values).

While paragraph [0057] of Rajan discusses estimating the distance of a user from the microphones, paragraph [0057] is completely silent and does not disclose or suggest that "the speech control unit being configured to **discriminate between sounds** originating from users who are located **in front of each other relative the microphone array**," as recited in independent claim 1, and similarly recited in independent claim 9. (Illustrative emphasis provided) Determining distances between users and microphones does not disclose or suggest to **discriminate between sounds** originating from users who are located **in front of each other relative the microphone array**. Further, FIG 1 of Rajan

shows three users sitting around a table, where none of the users are located **in front of each other relative the microphone array**.

Even, assuming arguendo, that Rajan discloses or suggests discriminating between sounds originating from users who are located in front of each other, and even if somehow Rajan discloses or suggests to discard some of the discriminated sounds, there is still no disclosure or suggestion in Rajan to discard an utterance originating from a **second** position **if not preceded by the recognition** of the predetermined keyword originating **from the second position**, as recited in independent claims 1 and 9.

In summary, it is respectfully submitted that Schroder, Kaufholz, Kataoka, and combination thereof, do not teach or suggest the present invention as recited in independent claim 1, and similarly recited in independent claim 9 which, amongst other patentable elements, recites (illustrative emphasis provided):

> the speech control unit being arranged to control the beam forming module, on basis of the recognition of the predetermined keyword, in order to enhance second components of the audio signals which represent a subsequent utterance originating from a second position of the user relative to the microphone array;
>
> wherein the recognition of the predetermined keyword at the second position calibrates the beam forming module to follow the user from the first position to the second position so that the subsequent utterance originating from the second position are accepted while utterances of other users at **other positions** are **discarded**, the second position including an orientation and a **distance relative to the microphone array**, and the speech control unit being configured to discriminate between **sounds** originating from users who are located **in front of each other relative the microphone array**; and
>
> wherein the subsequent utterance originating from the second position will be **discarded if not preceded** by the recognition of the

predetermined keyword originating **from the second position**.

Discriminate between sounds originating from users who are located in front of each other **relative the microphone array**, accepting a subsequent utterances originating from the second position while discarding utterances of other users at other positions, and discarding the subsequent utterance originating from the second position if not preceded by recognition of the predetermined keyword originating **from the second position**, are nowhere disclosed or suggested in Schroder, Kaufholz and Rajan, alone or in combination. Kataoka is cited to allegedly show other features and does not remedy the deficiencies in Schroder, Kaufholz and Rajan.

Accordingly, it is respectfully submitted that independent claims 1 and 9 are allowable. In addition, it is respectfully submitted that claims 2-8, 10, 15 and 20 are also allowable at least based on their dependence from independent claims 1 and 9.

Further, it is respectfully submitted that "the keyword recognition system is connected to one microphone of the microphone array for detecting the predetermined keyword, the keyword recognition system being further connected to the beam forming module for providing the detected predetermined keyword to the beam forming module," as recited in claim 20 is nowhere disclosed or suggested in Schroder, Kaufholz, Rajan, and combination thereof. On page 8 of the Final Office Action, it is alleged that FIG 2-3 and column 4, line 58 to column 5, line 40 of Kaufholz. This allegation is respectfully traversed.

A careful inspection of the noted sections of Kaufholz, such as FIG 2, clearly shows that a beam former 240 receives outputs of two microphones 204, 214 of an audio set 200 and a TV 210, where the beam former 240 output is connected to an audio cancellation module 100. FIG 2 shown NO keyword recognition system. However, FIG 3 shows a speech recognizer 300 which receives the output of an audio cancellation module 100 where, in the embodiment shown in FIG 3, the audio cancellation module 100 receives signals from microphones of the audio set 200 and the TV 210. The speech recognizer 300 is connected to a controller 310. FIGs 2 and 3 of Kaufholz do not disclose or suggest any keyword recognition system which is connected both to a microphone and to a beam forming module for providing the detected predetermined keyword to the beam forming module, as recited in claim 20. Rather, the speech recognizer 300 shown in FIG 3 of Kaufholz is merely connected between the audio cancellation module 100 and the controller 100. Accordingly, it is respectfully submitted that dependent claim 20 is allowable.

Claims 11-14 and 16-19 are said to be unpatentable over Schroder, Kaufholz, Rajan and Kataoka.

It is respectfully submitted that claims 11-14 and 16-19 should be allowed at least based on their dependence from independent claims 1 and 9.

In addition, Appellant denies any statement, position or averment of the Examiner that is not specifically addressed by the foregoing argument and response. Any rejections and/or points of argument not addressed would appear to be moot in view of the presented remarks. However, the Appellant reserves the right to submit further arguments in support of the above stated position, should that become necessary. No arguments are waived and none of the Examiner's statements are conceded.

CONCLUSION

Claims 1-20 are patentable over Schroder, Kaufholz, Rajan and Kataoka.

Thus, the Examiner's rejections of claims 1-20 should be reversed.

Respectfully submitted,

By_____
Dicran Halajian, Reg. 39,703
Attorney for Appellant
July 12, 2010

THORNE & HALAJIAN, LLP
Applied Technology Center
111 West Main Street
Bay Shore, NY 11706
Tel: (631) 665-5139
Fax: (631) 665-5101

CLAIMS APPENDIX

1.(Previously Presented) A speech control unit for controlling an apparatus on basis of speech, comprising:

a microphone array, comprising multiple microphones for receiving respective audio signals;

a beam forming module for extracting a speech signal of a user, from the audio signals as received by the microphones, by means of enhancing first components of the audio signals which represent an utterance originating from a first position of the user relative to the microphone array;

a speech recognition unit for creating an instruction for the apparatus based on recognized speech items of the speech signal; and

a keyword recognition system for recognition of a predetermined keyword that is spoken by the user and which is represented by a particular audio signal;

the speech control unit being arranged to control the beam forming module, on basis of the recognition of the predetermined keyword, in order to enhance second components of the audio signals which represent a subsequent utterance originating from a second position of the user relative to the microphone array;

wherein the recognition of the predetermined keyword at the second position calibrates the beam forming module to follow the user from the first position to the second

position so that the subsequent utterance originating from the second position are accepted while utterances of other users at other positions are discarded, the second position including an orientation and a distance relative to the microphone array, and the speech control unit being configured to discriminate between sounds originating from users who are located in front of each other relative the microphone array; and

wherein the subsequent utterance originating from the second position will be discarded if not preceded by the recognition of the predetermined keyword originating from the second position.

2.(Previously Presented) The speech control unit as claimed in claim 1, wherein the keyword recognition system is arranged to recognize the predetermined keyword that is spoken by another user and the speech control unit being arranged to control the beam forming module, on basis of this recognition, in order to enhance third components of the audio signals which represent another utterance originating from a third position of the other user relative to the microphone array.

3.(Previously Presented) The speech control unit as claimed in claim 1, wherein a first one of the microphones of the microphone array is arranged to provide the particular audio signal to the keyword recognition system.

4.(Previously Presented) The speech control unit as claimed in claim 1, wherein the beam forming module is arranged to determine a first position of the user relative to the microphone array.

5.(Original) An apparatus comprising:

a speech control unit for controlling the apparatus on basis of speech as claimed in claim 1; and

processing means for execution of the instruction being created by the speech control unit.

6.(Previously Presented) The apparatus as claimed in claim 5, the apparatus arranged to show that the predetermined keyword has been recognized.

7.(Previously Presented) The apparatus as claimed in claim 6, further comprising audio generating means for generating an audio signal in order to show that the predetermined keyword has been recognized.

8.(Original) A consumer electronics system comprising the apparatus as claimed in claim 5.

9.(Previously Presented) A method of controlling an apparatus on basis of speech, comprising the acts of:

receiving respective audio signals by means of a microphone array, comprising multiple microphones;

extracting a speech signal of a user, from the audio signals as received by the microphones, by means of enhancing first components of the audio signals which represent an utterance originating from a first position of the user relative to the microphone array;

recognizing a predetermined keyword that is spoken by the user based on a particular audio signal and controlling the extraction of the speech signal of the user, on basis of the recognition of the predetermined keyword, in order to enhance second components of the audio signals which represent a subsequent utterance originating from a second position of the user relative to the microphone array while discarding utterances of other users at other positions, the second position including an orientation and a distance relative to the microphone array so that sounds originating from users who are located in front of each other relative the microphone array are discriminated;

creating an instruction for the apparatus based on recognized speech items of the speech signal; and

discarding the subsequent utterance originating from the second position if not preceded by the recognition of the predetermined keyword originating from the second position.

10.(Previously Presented) The speech control unit of claim 1, wherein the user is informed by indications that the speech control unit is not active, is in an active state and ready to receive the utterance, or is in a state of calibration.

11.(Previously Presented) The speech control unit of claim 10, wherein the indications include an animal in a sleeping state indicating that the speech control unit is not active, and in an awake state indicating that the speech control unit is in the active state.

12.(Previously Presented) The speech control unit of claim 11, wherein progress of the active state is indicated by an angle of ears of the animal.

13.(Previously Presented) The speech control unit of claim 12, wherein the ears are fully raised at a beginning of the active state, and fully down at an end of the active state.

14.(Previously Presented) The speech control unit of claim 11, wherein the animal has an understanding look when the utterance is recognized and a puzzled look when the utterance is not recognized.

15.(Previously Presented) The method of claim 9, further comprising the act of informing the user by indications that the apparatus is not active, is in an active state and ready to receive the utterance, or is in a state of calibration.

16.(Previously Presented) The method of claim 15, wherein the indications include an animal in a sleeping state indicating that the speech control unit is not active, and in an awake state indicating that the speech control unit is in the active state.

17.(Previously Presented) The method of claim 16, wherein progress of the active state is indicated by an angle of ears of the animal.

18.(Previously Presented) The method of claim 17, wherein the ears are fully raised at a beginning of the active state, and fully down at an end of the active state.

19.(Previously Presented) The method of claim 16, wherein the animal has an understanding look when the utterance is recognized and a puzzled look when the utterance is not recognized.

20.(Previously Presented) The speech control unit of claim 1, wherein the beam forming module is connected to the microphone array, and the keyword recognition system

is connected to one microphone of the microphone array for detecting the predetermined keyword, the keyword recognition system being further connected to the beam forming module for providing the detected predetermined keyword to the beam forming module.

EVIDENCE APPENDIX

None

RELATED PROCEEDINGS APPENDIX

None